Data Supplement S1. Supplemental Material: "The Quick Walk Test: A Non-invasive Test to Assess the Risk of Mechanical Ventilation during COVID-19 Outbreaks"

Figure S1. Flow-chart describing the patient selection and classification using three QWT criteria.

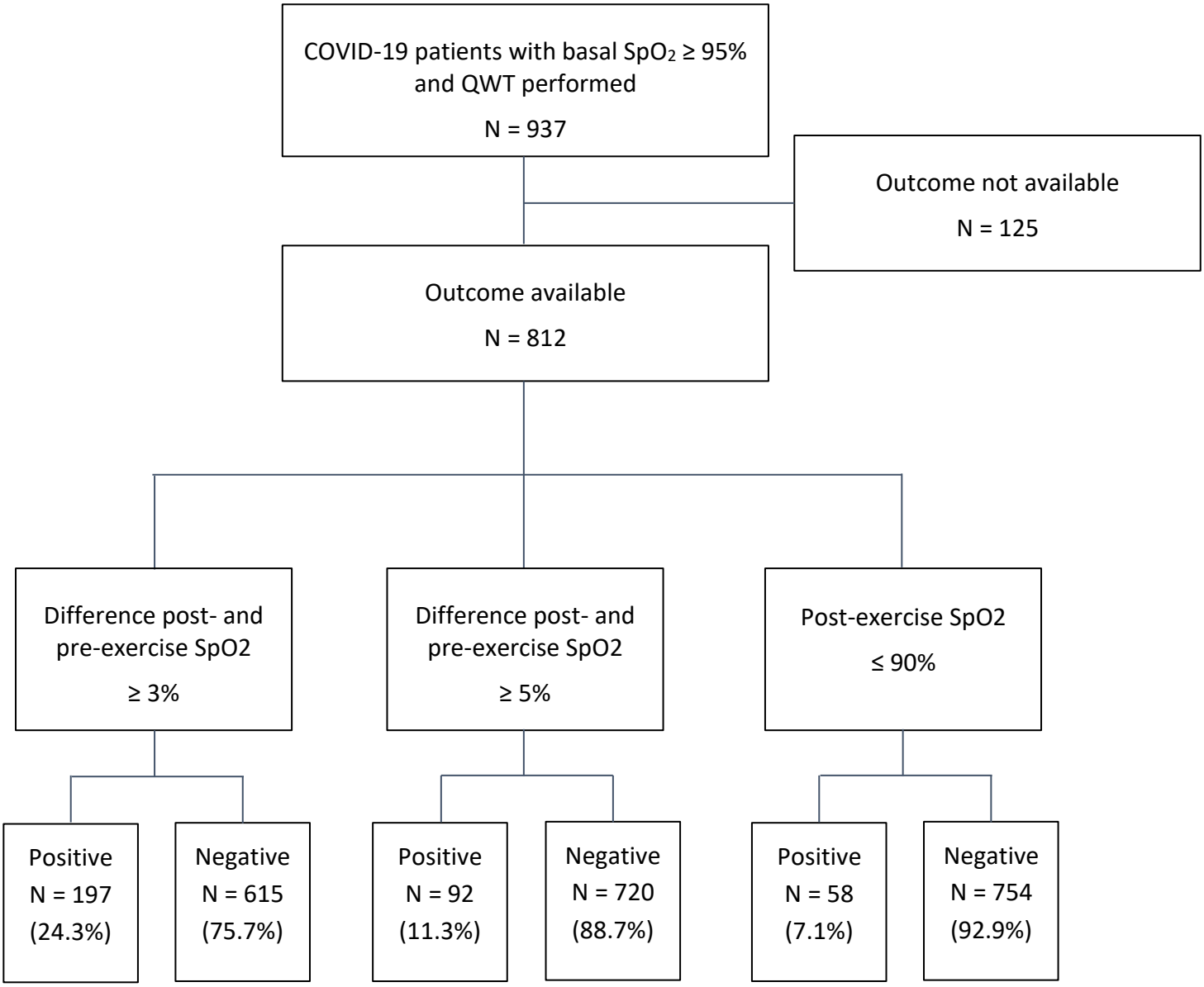


Table S1. Demographic and clinical characteristics of the enrolled patients.

Variables	Enrolled Patients (N = 812)
Age	
Mean (SD)	50.1 (14.3)
Median (Q1-Q3)	50.5 (40-59)
Sex (Female) – N (%)	393 (48.4)
Vital Signs at ED presentation:	
Heart Rate (beats per minute) – N (%)	
<70	53 (7.0)
70 – 99	444 (59.0)
>=100	256 (34.0)
Missing	59
Respiratory Rate (breaths per minute) – N (%)	
<=22	625 (92.9)
>22	48 (7.1)
Missing	139
Pre-exercise SpO_2 (%) – N (%)	
95 – 96	185 (22.8)
97 – 98	416 (51.2)
99 – 100	211 (26.0)
Systolic Blood Pressure (mmHg) – N (%)	
<100	7 (1.0)
100 – 139	459 (64.9)
>=140	241 (34.1)
Missing	105
Diastolic Blood Pressure (mmHg) – N (%)	
<60	23 (3.3)
60 – 89	510 (72.1)
>=90	174 (24.6)
Missing	105
Temperature (°C) – N (%)	
<37.5	533 (79.3)
37.5 – 38.4	114 (17.0)
>=38.5	25 (3.7)
Missing	140

S1. Robustness of the study results to the exclusion of the patients with missing outcome.

To evaluate the sensitivity of the study results to the exclusion of the patients with missing outcomes, we provide a description of the excluded patients and the results of two sensitivity analyses.

Description of the Excluded Patients

All the 937 patients included in our study sample performed the QWT. However, the outcome (invasive mechanical ventilation) was missing for 125 patients (13.3%). To assess the performance of the Quick Walk Test (QWT), the patients with missing outcome were excluded from our main analysis.

A description of the excluded patients is provided in Table S2. Overall, the distributions of the characteristics are comparable to those of patients whose outcome was collected (Table S1). Limited discrepancies between the distributions were observed in age, respiratory rate and pre-exercise peripheral oxygen saturation. With respect to these characteristics, excluded patients appear to be slightly younger and marginally less severe.

Table S3 provides the proportion of excluded patients with positive QWT, for each of the three interpretations of the QWT. Notably, the proportions of positive tests are similar to those observed in the patients included in the main analysis (see Table 1 of the manuscript). These comparisons suggest that the excluded patients were not considerably different, with respect to the observed characteristics, from those included in the main analysis.

Sensitivity Analysis 1: Test Ignorance Regions

We performed a sensitivity analysis by constructing the test ignorance regions for the three interpretations of the QWT, using the methodology described by Kosinski and Barnhart[1]. The test ignorance region is the set of all sensitivity and the specificity values that are consistent with the study data. When the reference standard of a diagnostic test (the outcome in our case) has missing values, such a region contains all the values of sensitivity and specificity that could have been estimated, considering all possible outcome values for the missing data. In particular, it is possible to derive upper and lower bounds for the measures of performance of a test.

Figure S2, S3 and S4 depict the test ignorance regions for the three interpretations of the test, considering as positive patients with a loss of saturation of 3 percentage points after QWT, a loss of saturation of 5 percentage points after QWT and a post-exercise saturation less than or equal to 90%, respectively. The limited height of the regions indicates the robustness of the estimates of the specificity, i.e., the proportion of patients with negative QWT among those who were not mechanically ventilated. In particular, whatever the outcome values of the 125 excluded patients were, the specificity of the third version of the test is bounded in a very high range of values, within the 90%-95% interval. Conversely, the width of the regions suggests that the sensitivity of the test (i.e., the proportion of positive QWT among the patients who were mechanically ventilated) may vary considerably depending on the value of the missing outcomes. This was expected, given the very small number of patients who were mechanically ventilated. Note that the test ignorance regions are constructed accounting for all possible outcome values in the patients with missing outcome, including very unrealistic scenarios where very high proportions of the excluded patients were mechanically ventilated. Unfortunately, this methodology does not account for the likelihood of the scenarios given what we observed, and the resulting plots should be consequently interpreted with care.

Sensitivity Analysis 2: Bayesian Estimation of the Performance of the Test

We performed a second sensitivity analysis using a Bayesian approach to estimate the performance of the QWT, using the methodology described by Pennello[2]. The author described two Bayesian models to estimate the performance of a diagnostic test (sensitivity, specificity and positive and negative predictive values) in studies where the reference standard may be missing for some patients. Both models depend on the probability of not observing the reference standard (i.e., the probability of outcome missingness in our case) but use different assumptions. In the first case, this probability is assumed to only depend on the result of the test, regardless of the outcome value (assumption 1). Conversely, in the second model, it is assumed to depend on the outcome value but not on the result of the test (assumption 2). Notably, the model where the probability of outcome missingness is assumed to depend both on the outcome and test value cannot be considered, as its parameters are not identifiable[2].

The methodology is designed to account for the missingness of the outcomes and incorporating it into the uncertainty of the estimates of model performance. As in all Bayesian models, a prior distribution is assumed for the key parameters and the collected data are used to derive posterior distributions, which provide estimates for the parameters of interest. We used uninformative priors on all the parameters, as suggested[2]. For the model using assumption 1, it is possible to specify the posterior distributions in closed form. This is not possible for the model using assumption 2, which is fit using a Gibbs sampling algorithm[2]. Posterior estimates are based on 10,000 samples, obtained after discarding a burn-in of 50,000 samples. Convergence was assessed by inspection of the trace plots and by applying the Heidelberger-Welch's and Geweke's convergence diagnostics[3] (data not shown).

Table S4 and S5 provide the estimates of sensitivity, specificity and positive and negative predictive values according to the models using assumption 1 and 2, respectively, for all of the three interpretations of the QWT. Notably, all the measures of performance where estimated similarly from the two models, suggesting that the assumption on the probability of outcome missingness does not heavily affect the result.

The estimated specificity, positive and negative predictive values were also similar to those of the main analysis (see the section Results of the manuscript). With respect to the sensitivity, the estimates from both the models are lower than what observed in the main analysis (for the interpretation of the test based on the post-exercise QWT, 75.3% and 70.7% vs. 83.3%). The reason for such a difference is due to the uninformative prior distributions, which assume the complete range of sensitivities (0 to 100%) as equally likely. As the patients who were mechanically ventilated were only 6, the estimate of the sensitivity is importantly affected by the prior distribution and, in particular, it is pulled toward lower values. Nevertheless, the very wide credible intervals of the estimates of the sensitivity confirmed the high degree of uncertainty on this measure of performance emerging from the main analysis.

In summary, the results of this sensitivity analysis are consistent with what emerged from the main analysis, where patients with missing outcomes where excluded.

Table S2. Characteristics of the 125 enrolled patients with missing outcome (invasive mechanical ventilation).

Variables	Description (N = 125)
Age	
Mean (SD)	44.0 (14.2)
Median (Q1-Q3)	45 (33-53)
Sex (Female) – N (%)	59 (47.2)
Vital Signs at ED presentation:	
Heart Rate (beats per minute) – N (%)	
<70	8 (7.1)
70 – 99	68 (60.2)
>=100	37 (32.7)
Missing	12
Respiratory Rate (breaths per minute) – N (%)	
<=22	109 (97.3)
>22	3 (2.7)
Missing	13
Pre-exercise SpO2 (%) – N (%)	
95 – 96	12 (9.6)
97 – 98	59 (47.2)
99 – 100	54 (43.2)
Systolic Blood Pressure (mmHg) – N (%)	
<100	1 (1.0)
100 – 139	70 (68.0)
>=140	32 (31.1)
Missing	22
Diastolic Blood Pressure (mmHg) – N (%)	
<60	2 (1.9)
60 – 89	72 (69.9)
>=90	29 (28.2)
Missing	22
Temperature (°C) – N (%)	
<37.5	99 (82.5)
37.5 – 38.4	17 (14.2)
>=38.5	4 (3.3)
Missing	5

Table S3. Interpretation of the QWT according to the different criteria in group of patients with missing outcome.

QWT Criteria	Patients with missing outcome N = 125
Loss of 3 percentage points after QWT - N (column %)	
Positive	27 (21.6)
Negative	98 (78.4)
Loss of 5 percentage points after QWT - N (column %)	
Positive	19 (15.2)
Negative	106 (84.8)
Post-QWT saturation ≤ 90% - N (column %)	
Positive	11 (8.8)
Negative	114 (91.2)

Table S4. Estimates of performance of the test (and 95% credible intervals) based on the Bayesian model that assumes the probability of outcome missingness to depend on the test result but not on the outcome value.

QWT Criteria and Measures of Performance	Estimate and 95% Credible Intervals
Loss of 3 percentage points after QWT	
Sensitivity	74.7% (42.2% - 96.2%)
Specificity	76.5% (73.8% - 79.2%)
Positive Predictive Value	3.0% (1.1% - 5.8%)
Negative Predictive Value	99.7% (99.1% - 100.0%)
Loss of 5 percentage points after QWT	
Sensitivity	75.7% (42.9% - 96.4%)
Specificity	88.7% (86.6% - 90.6%)
Positive Predictive Value	6.4% (2.4% - 12.1%)
Negative Predictive Value	99.7% (99.2% - 100.0%)
Post-QWT saturation ≤ 90%	
Sensitivity	75.3% (43.1% - 96.2%)
Specificity	93.2% (91.5% - 94.7%)
Positive Predictive Value	10.0% (3.8% - 18.7%)
Negative Predictive Value	99.7% (99.3% - 100.0%)

Table S5. Estimates of performance of the test (and 95% credible intervals) based on the Bayesian model that assumes the probability of outcome missingness to depend on the outcome value but not on the result of the test.

QWT Criteria and Measures of Performance	Result and 95% Credible Intervals
Loss of 3 percentage points after QWT	
Sensitivity	68.4% (30.1% - 95.3%)
Specificity	76.7% (73.9% - 79.4%)
Positive Predictive Value	4.5% (1.4% - 9.7%)
Negative Predictive Value	99.1% (96.1% - 99.9%)
Loss of 5 percentage points after QWT - N (column %)	
Sensitivity	69.9% (31.6% - 95.6%)
Specificity	89.2% (87.1% - 91.2%)
Positive Predictive Value	11.6% (4.0% - 21.0%)
Negative Predictive Value	99.1% (96.3% - 99.9%)
Post-QWT saturation ≤ 90% - N (column %)	
Sensitivity	70.7% (34.6% - 95.8%)
Specificity	93.5% (91.8% - 95.0%)
Positive Predictive Value	13.7% (5.1% - 25.4%)
Negative Predictive Value	99.5% (98.1% - 100.0%)

Figure S2. Test ignorance region for the interpretation of the test considering as positive the patients with a loss of saturation of 3 percentage points after QWT.

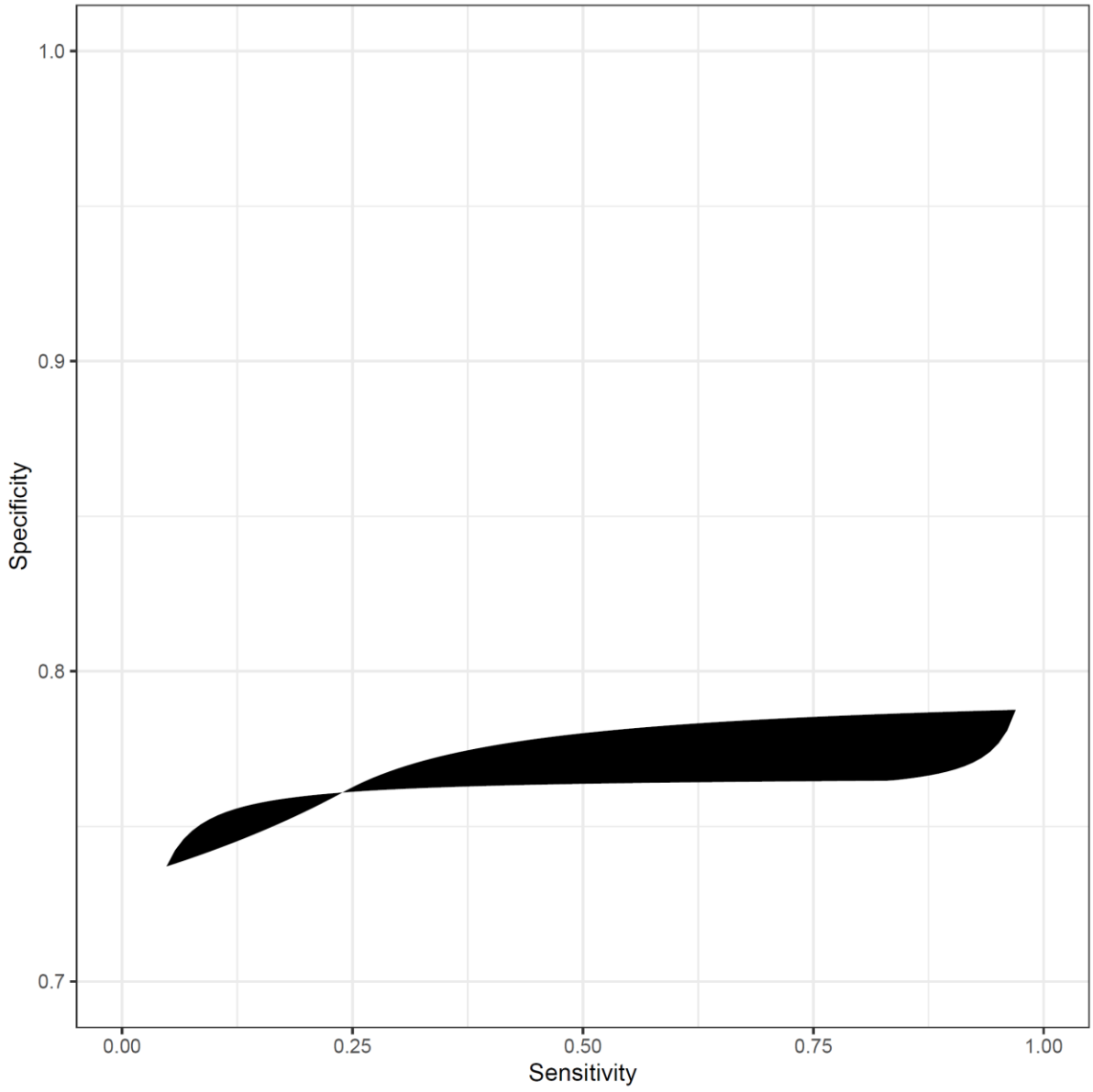


Figure S3. Test ignorance region for the interpretation of the test considering as positive the patients with a loss of saturation of 5 percentage points after QWT.

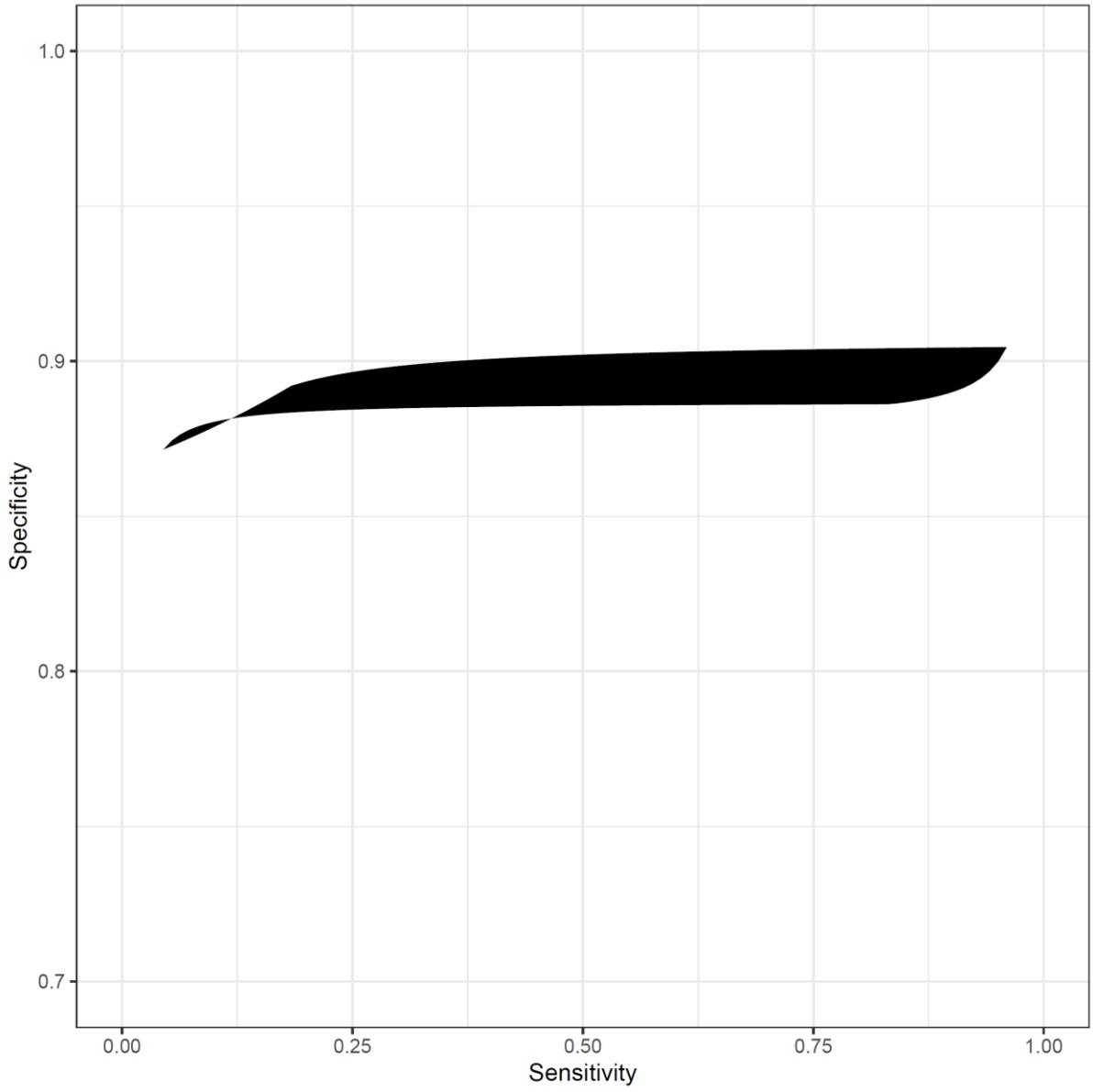
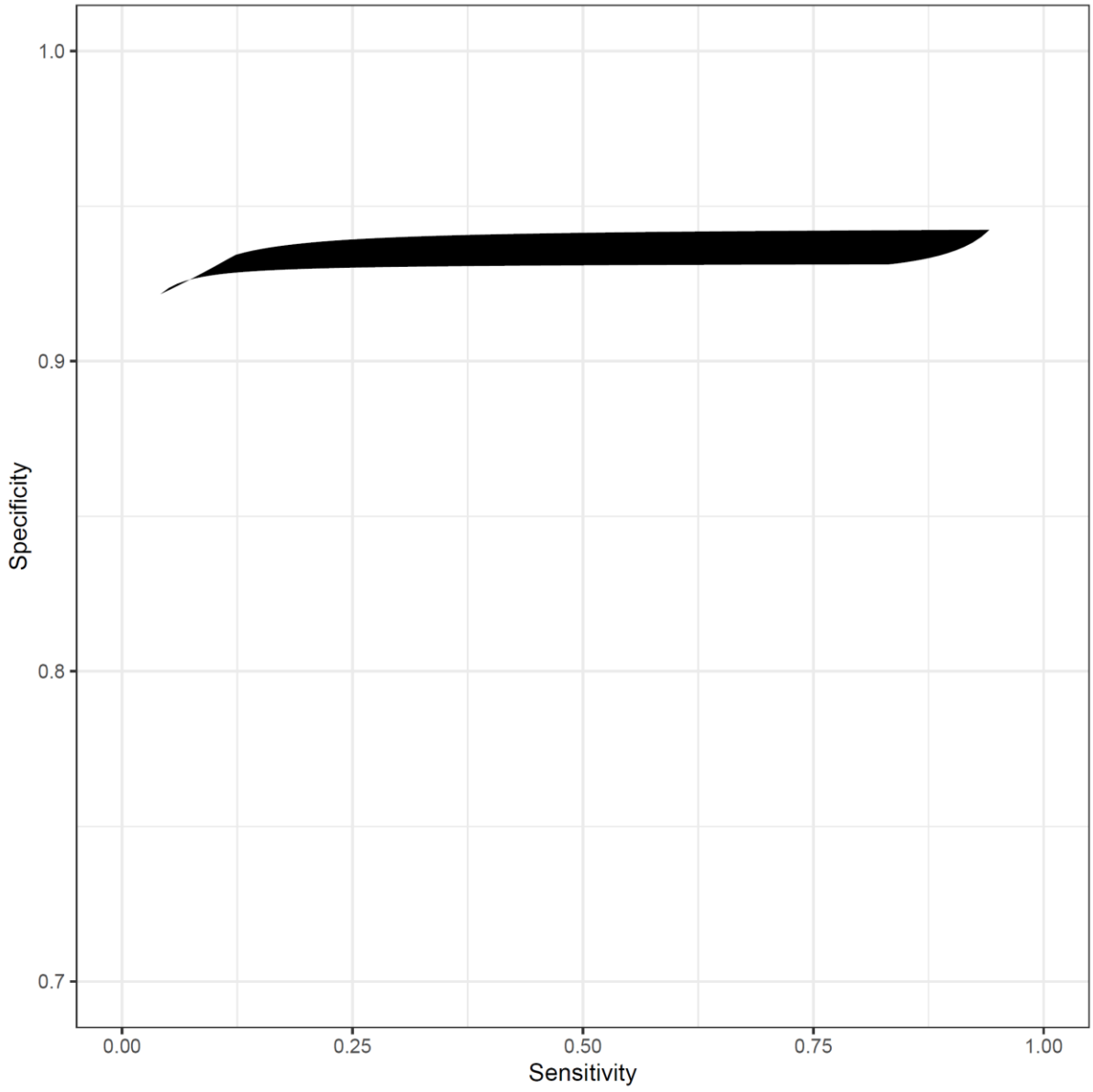


Figure S4. Test ignorance region for the interpretation of the test considering as positive the patients with a post-exercise saturation less than or equal to 90%.



References

1. Kosinski AS, Barnhart HX. A global sensitivity analysis of performance of a medical diagnostic test when verification bias is present. *Statistics in Medicine*. 2003;22(17):2711-2721. doi:10.1002/sim.1517
2. Pennello GA. Bayesian analysis of diagnostic test accuracy when disease state is unverified for some subjects. *J Biopharm Stat*. 2011;21(5):954-970. doi:10.1080/10543406.2011.590921
3. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis, Third Edition*. CRC Press; 2013.